

Modelado Semántico y Centrado en el Usuario de Servicios Adaptados al Contexto de Uso

Juan Carlos Yelmo García Yod Samuel Martín García Beatriz San Miguel González
Departamento de Ingeniería de Sistemas Telemáticos – Universidad Politécnica de Madrid
E.T.S.I. Telecomunicación – Av. Complutense, 30 – “Ciudad Universitaria”
28040 Madrid - España

(+34) 91 336 68 30
jcyelmo@dit.upm.es

(+34) 91 336 73 66 x 3028
samuelm@dit.upm.es

(+34) 91 336 73 66 x 3028
smiguel@dit.upm.es

RESUMEN

Cuando los usuarios acceden a servicios telemáticos, esperan obtener experiencias cada vez más adaptadas a su contexto específico, en especial, ante la gran cantidad de contenidos en línea a su alcance a través de dichos servicios. Para que estos servicios proporcionen funcionalidades centradas en el usuario – por ejemplo, búsquedas relevantes, adaptación de contenidos, personalización y recomendación– es necesario que los contenidos estén anotados semánticamente y disponer de un modelo preciso que represente al usuario y su contexto de uso. En este escenario, presentamos un marco genérico de servicios centrados en el usuario basado en la caracterización semántica de los usuarios y su mapeo con los contenidos. En concreto, aplicamos dicho marco a un buscador semántico que caracteriza automáticamente tanto los contenidos a los que los usuarios pueden acceder como el contexto de uso desde el que se accede. La solución propuesta incluye modelos de usuario dinámicos y adaptativos, con información explícita e implícita; así como descriptores de los contenidos que pueden utilizarse para determinar cuáles son más adecuados para cada usuario.

Categories and Subject Descriptors

H.3.5 [Information Storage and Retrieval]: Online Information Services – *Web-based services*.

General Terms

Design, Human Factors

Palabras clave

Modelo de usuario, metadatos, descripción de contenidos, adaptación, buscador semántico, personalización

1. INTRODUCCIÓN

Hoy en día, los usuarios de servicios telemáticos están acostumbrados a un acceso ubicuo, con disponibilidad permanente e independiente del dispositivo. En este contexto, los

usuarios esperan obtener una experiencia adaptada a sus preferencias personales y necesidades –lo cual varía según su contexto de uso o las tareas que realizan en un determinado momento–. Para dar respuesta a esta demanda, surgieron las herramientas centradas en el usuario, que adaptan sus funcionalidades a cada usuario y les permiten controlar el proceso, a la vez que reducen la carga cognitiva asociada al mínimo.

El contenido de esta ponencia se enmarca dentro de un proyecto en el que se desarrolla un buscador semántico multimedia [3]. Presentamos aquí el trabajo en curso dentro de dicho proyecto, cuyo objetivo es el de ofrecer un marco genérico para otros servicios, con el fin de mejorar las funcionalidades centradas en el usuario. Concretamente, utilizando este marco dentro de nuestro servicio de búsquedas, ofrecemos nuevas operaciones denominadas:

- *Búsqueda centrada en el usuario*: definida ésta como la selección de un subconjunto limitado de contenidos pertenecientes a un repositorio, basándose esta selección en la petición de búsqueda del usuario (estructurada o no estructurada). Gracias a los datos recuperados del modelo del usuario que efectúa la consulta, se mejora la precisión de la búsqueda, su relevancia para el usuario y su corrección.
- *Recomendación*: misma definición que la operación de *búsqueda*, pero sin partir de la intervención del usuario como en el caso anterior. La recomendación funciona en modo *push*, lanzada por el servicio sin recibir ninguna otra entrada.
- *Personalización*: selección de un contenido multimedia de entre un pequeño conjunto de contenidos equivalentes (por ejemplo, varias versiones del mismo contenido).
- *Adaptación*: transformación de un determinado contenido multimedia recibido como entrada. Ésta es la única operación que no ofrece un conjunto de enlaces como resultado, sino que implica la entrega directa del contenido al usuario.

En cada una de las operaciones presentadas, los fundamentos de las funcionalidades centradas en el usuario son los mismos. El servicio se apoya tanto en descripciones enriquecidas y semánticas de los contenidos incluidos en un repositorio (distribuido o centralizado) como en el modelo de los usuarios que acceden a dichos contenidos. Gracias a la existencia de las

descripciones de ambos elementos (contenidos y usuarios), se pueden realizar operaciones de mapeo que permiten seleccionar el mejor contenido disponible para un usuario específico. Consecuentemente, esto implica haber obtenido con anterioridad un conocimiento estructurado de los usuarios (que represente sus necesidades y preferencias) y de los contenidos (para valorar cómo éstos casan con un usuario específico) [7]. En concreto, proponemos un marco semántico que captura, representa y almacena estos dos dominios, para que así el servicio de búsqueda pueda proveer funcionalidades centradas en el usuario. Otros servicios de búsqueda personalizados [4] aplican técnicas de inferencia a la información de la que disponen sobre el usuario para añadir nuevos términos a la consulta más allá de los que el usuario incluye explícitamente. Por el contrario, nuestro enfoque realiza la correspondencia entre el modelo del usuario y los contenidos incluyendo descriptores que no se encuentran entre los términos del contenido sino que expresan parámetros sintéticos (accesibilidad, legibilidad, lenguaje, etc.) resultado de distintos análisis.

Este marco presentado constituye un servicio auxiliar al de búsqueda, proporcionando un repositorio de metadatos de contenidos y usuarios que puede utilizarse en la implementación de funcionalidades centradas en el usuario. Aunque presentamos la aplicación del marco para un servicio de búsquedas, otros servicios centrados en el usuario, podrían utilizar también este enfoque. La Figura 1 resume el escenario propuesto para el servicio de búsquedas. Entre los elementos que participan en el escenario, se incluye: a) un motor de búsquedas enriquecido con funcionalidades de mapeo de contenidos a usuarios, b) herramientas de anotación de contenidos que se apoyan en información explícita e implícita y c) un repositorio de modelos de usuario que la almacena. Los usuarios siempre desempeñan un rol activo respecto a cada uno de los componentes.

2. MODELADO DEL USUARIO Y DEL CONTEXTO DE USO

Para que el servicio pueda incorporar funcionalidades centradas en el usuario, el servicio debe disponer de un modelo de cada uno de los usuarios que acceden a él [9]. Un modelo de usuario consiste en una *descripción computacional de un determinado usuario en un sistema que incluye información relevante del usuario, ofreciendo conocimientos de manera estructurada sobre sus actividades, preferencias y capacidades*. Para obtener una imagen real de los usuarios, el modelo de usuario que proponemos no sólo incluye información específica de éstos, sino que además cubre el resto de elementos relativos a su contexto de uso: conocimiento, habilidades, objetivos, cultura, preferencias, comportamiento, destrezas y capacidades sensoriales, organización, entorno, plataforma software y hardware utilizada, etc.

A continuación, describimos cómo se obtiene esta información y presentamos el modelo de usuario que hemos considerado dentro de nuestro marco, junto con los retos de investigación que implica.

2.1 Obtención de información del usuario

Distinguimos dos tipos de información de usuario respecto a la forma en la que ésta se captura:

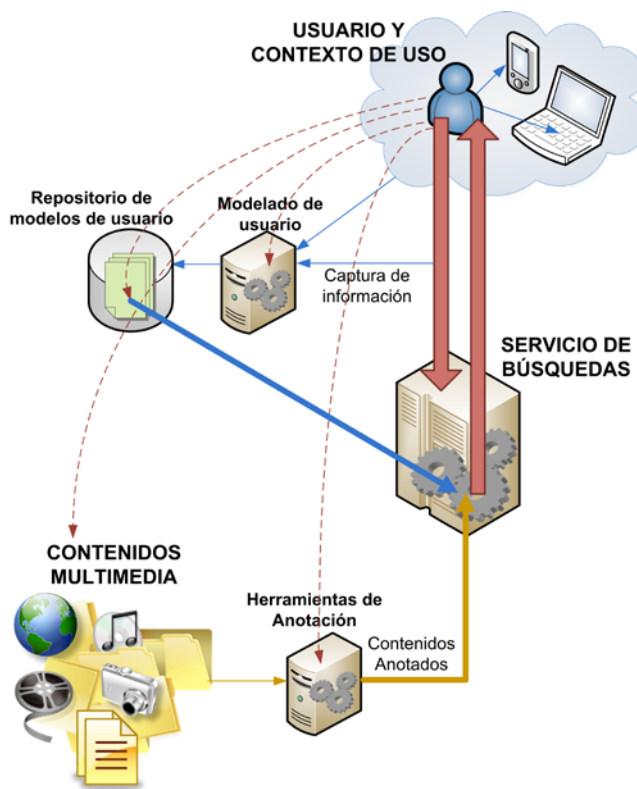


Figura 1. Marco genérico para ofrecer funcionalidades centradas en el usuario aplicado al servicio de búsquedas.

- *Información explícita:* los usuarios introducen personalmente esta información a través de formularios o encuestas. La información contenida puede incluir información demográfica, valoraciones numéricas, opiniones escritas, etc.
- *Información implícita:* se extrae a partir de la actividad o interacción del usuario con el servicio, sin intervención directa por su parte. Utilizando técnicas de minería de datos, la información implícita relacionada con las preferencias del usuario puede inferirse a partir del historial de navegación del usuario o de registros de interacción previamente almacenados.

Parece obvio pensar que la información obtenida directamente a través de los usuarios será más precisa y veraz que la información que puede inferirse a partir de los datos capturados de las interacciones del usuario con el servicio. Sin embargo, hay que considerar que los usuarios no siempre están dispuestos a colaborar rellenando formularios o encuestas, y que pueden introducir datos incorrectos, dada su falta de interés, o las molestias que el proceso puede implicar. Este problema radica en la propia esencia de la información explícita: la exigencia de contribuir dicha información puede suponer una sobrecarga para los usuarios. Para evitar esto, es necesario que los usuarios encuentren motivaciones que les hagan colaborar y aportar información apropiada –numerosos servicios demuestran que los usuarios se acaban involucrando si reciben los incentivos adecuados–. Consideramos que se les puede motivar a contribuir mejorando su experiencia de usuario y otorgándoles beneficios palpables, mediante el apoyo en funcionalidades centradas en el

usuario. Además, si este proceso es transparente para los usuarios y pueden controlar los datos que se almacenan sobre ellos, inferimos que mantendrán y actualizarán voluntariamente su información. En este punto, nuestro objetivo es proporcionar interfaces simples y sencillas que permitan la captura de la información adecuada, sin sobrecargar a los usuarios con múltiples preguntas; así como ofrecer incentivos que motiven a los usuarios para tomar parte activa en el control de su propio modelo.

2.2 Definición del modelo de usuario

Proponemos un modelo de usuario dinámico a partir de la fusión de información explícita e implícita de los usuarios, donde los usuarios puedan controlar la creación y evolución de su propio modelo. Encontramos, pues, imprescindible que los usuarios sean modelados como entidades únicas e individuales, ya que la diferencia entre ellos es lo suficientemente relevante en servicios telemáticos centrados en el usuario, que deben adaptarse a las características específicas de cada uno de ellos. Esto requiere el uso de modelos dinámicos de usuario, ya que nos permiten variar la información a lo largo del tiempo, pudiendo modificarse, ampliarse o mejorarse determinada información, para adaptarse a las circunstancias —y por tanto, a diferentes usuarios—. En definitiva, la principal ventaja de estos modelos dinámicos es este último punto: el modelo de usuario propuesto permite caracterizar a los usuarios individualmente, ya que se adapta a diferentes usuarios [5]. Anticipamos algunos de los enfoques que aplicaremos con este fin, basados en inferencias probabilísticas de las características del usuario, por ejemplo: redes bayesianas, agrupación de usuarios en estereotipos, filtrado colaborativo, etc.

Por otra parte, según se ha mencionado, queremos otorgar al usuario el control sobre su propio modelo de usuario. En este sentido, basándonos en los trabajos de Ahn et al.[1], incorporamos las interfaces necesarias que permiten presentar la información almacenada, en lo que se puede denominar un *modelo abierto de usuario*. Este modelo abierto incluye mecanismos que permiten a los usuarios controlar, modificar y borrar su información. Es importante señalar que el usuario no simplemente controla la información almacenada explícitamente por él mismo, sino también la información inferida por el servicio. Gracias a la modificación que los usuarios pueden realizar de la información almacenada, se obtiene una realimentación muy provechosa que puede utilizarse para mejorar las técnicas de inferencia utilizadas. Además, si aparecen conflictos o incoherencias entre la información explícita e implícita del usuario, se pueden resolver a través de modelos probabilísticos; aunque en nuestro marco, el usuario siempre tendrá la última palabra sobre la información que se ha almacenado sobre él.

Los usuarios pueden beneficiarse también del modelo abierto de usuario, ya que pueden controlar y salvaguardar su privacidad: vigilando, modificando y borrando sus datos personales de una forma clara y transparente. Gracias a esta transparencia, asumimos que los usuarios serán más permisivos a la hora de consentir la recogida de información por parte del servicio.

3. ANOTACIÓN DE CONTENIDOS

Hemos considerado tres cuestiones relacionadas con la anotación de contenidos: cuál es su alcance, cómo se generan y qué estructura se usa para representarlos.

3.1 Alcance

El alcance de las anotaciones se refiere a la determinación de qué descriptores se deben tener en cuenta. En concreto, queremos obtener descriptores o anotaciones que: (a) sean relevantes para las operaciones centradas en el usuario presentadas en la sección 1, y (b) impliquen restricciones significativas sobre los posibles contextos de uso desde los que se accede a los servicios. En este sentido, consideramos tres tipos de descriptores:

- *Etiquetas semánticas*: tema tratado en el contenido, dominio del contexto, entidades relevantes que aparecen en el contenido, etc.
- *Contexto*: fecha y localización representados en el contenido.
- *Características de presentación*: formato del contenidos, lenguaje natural y modalidades de presentación.

3.2 Generación

Para generar la descripción de los contenidos y su consiguiente anotación, existe la opción directa de analizar los contenidos cada vez que se accede a éstos. Sin embargo, esto es totalmente irrealizable por la gran carga computacional asociada. En su lugar, utilizamos tres enfoques para anotar contenidos:

- *Anotación (manual o asistida) por parte de los productores de contenidos*. Idealmente, durante el proceso de producción, el autor u otro miembro del equipo de producción pueden incluir anotaciones semánticas en los contenidos. Sin embargo, esto no siempre es factible; especialmente cuando el contenido está creado por prosumidores (*prosumers*) o usuarios no profesionales [8].
- *Enriquecimiento del contenido por parte de la comunidad*. Las anotaciones o metadatos pueden asociarse a los contenidos después de que éstos hayan sido creados. Esto permite aprovechar la sabiduría colectiva para obtener anotaciones de los contenidos generadas por los usuarios. Ejemplos de esta práctica son las *folksonomías* [6] (que etiquetan contenidos con etiquetas semánticas) y la accesibilidad social [10] (donde se enriquecen los contenidos con indicaciones para las ayudas técnicas).
- *Generación automática y almacenamiento de anotaciones mediante herramientas de análisis*. El mecanismo menos costoso para añadir anotaciones consiste en la automatización del proceso. Realizamos este proceso a través de herramientas externas de fuente abierta como GATE (General Architecture for Text Engineering) y Apache Tika, u otras propias de socios del proyecto, que permiten tanto extraer metadatos incrustados como inferir nuevo conocimiento a partir del texto. Mediante el desarrollo de nuevos módulos para dichas herramientas ejecutamos distintos procesos: identificación del idioma, análisis morfológico, reconocimiento de entidades, evaluación heurística de accesibilidad, medida de legibilidad, etc.

3.3 Representación

A la hora de decidir cómo se deben almacenar y transmitir los metadatos o anotaciones, consideramos que las descripciones de los contenidos anotados deben seguir un formato estándar que permita fácilmente su reutilización en cualquier herramienta, con independencia del contenido al que se refiera. Hemos considerado varias alternativas complementarias:

- *Modelos genéricos e independientes del vocabulario*, como el Protocolo Web de Descripción de Recursos [2] (POWDER – Protocol for Web Description Resources) que permite una sencilla integración de información de distintos dominios (accesibilidad, contenido adulto, acceso móvil, etc.)
- Estándares de *descripción de contenidos multimedia* como MPEG-7.
- Modelos de representación de *información bibliográfica* como Dublin Core, Machine Reading Cataloging (MARC), Metadata Object Description Schema (MODS) o la Text Encoding Initiative (TEI).
- Formatos de metadatos orientados a contenidos multimedia de noticias como Information Interchange Module y NewsML.

Es importante señalar que todos estos formatos se pueden integrar fácilmente en un marco genérico compatible con diferentes modelos (como el mencionado POWDER). Esto se debe a que todos ellos son vocabularios controlados para los que hay definida una representación en XML. Además, esto facilita almacenar las anotaciones dentro de la ontología multilingüe, multidominio y multimedia utilizada en el proyecto.

4. CONCLUSIONES Y TRABAJO FUTURO

A lo largo de la ponencia hemos visto cómo un servicio telemático centrado en el usuario puede aprovechar la caracterización de los contenidos y de los usuarios para mejorar sus resultados. Hemos presentado un trabajo en curso que define un marco genérico, realizado en el contexto de un servicio telemático de búsquedas semánticas, pero que puede extenderse a otros servicios para ofrecer funcionalidades centradas en el usuario. A su vez, el marco también puede caracterizarse como *centrado en el usuario*, no sólo por permitir que los usuarios accedan de manera personalizada, sino porque también les otorga el control del proceso de anotación, dejándoles decidir sobre los siguientes aspectos:

- La anotación de los contenidos, que puede realizarse por el creador en tiempo de producción, o estar automatizada, o incluso realizarse comunitariamente por varios usuarios en entornos colaborativos.
- Los modelos de los usuarios, que pueden crearse sólo con información inferida pero también pueden depender de la información provista por el usuario. Además, los usuarios a quienes les inquiete su privacidad siempre podrán delimitar y especificar qué información se almacena sobre ellos en cada momento.

El trabajo presentado continúa en proceso de desarrollo, y se prevé implementar el marco de anotación con la integración de herramientas externas de análisis de contenidos y de análisis de registros de uso para, finalmente, integrarlo dentro del servicio de búsqueda.

5. AGRADECIMIENTOS

El trabajo presentado forma parte del proyecto CENIT Buscamedia con referencia CEN-20091026, financiado parcialmente por el Centro para el Desarrollo Tecnológico e Industrial (CDTI) del Ministerio de Ciencia e Innovación de España. Los autores quieren agradecer la colaboración del resto de los socios (www.cenitbuscamedia.es). Específicamente, esta parte del proyecto ha sido contratado por GFI Informática a la Universidad Politécnica de Madrid.

6. REFERENCIAS

- [1] Ahn, J., Brusilovsky, P., Grady, J., He, D., and Syn, S. Y. Open user profiles for adaptive news systems: help or harm?. In Proceedings of the 16th international Conference on World Wide Web (Banff, Alberta, Canada, May 08 - 12, 2007). WWW '07, pp. 11-20. ACM, New York, NY (2007).
- [2] Archer, P., Smith, K., Perego, A. (eds.), Protocol for Web Description Resources (POWDER): Description Resources, W3C Recommendation. W3C (2009). <http://www.w3.org/TR/powder-dr/>
- [3] Cenit Buscamedia, Página del proyecto, <http://www.cenitbuscamedia.es/>
- [4] Cui H., Wen J., Nie J., and Ma W. Probabilistic query expansion using query logs. In WWW '02: Proc of the 11th Int'l conf on WWW, pages 325–332, 2002.
- [5] Gaudioso, E., Boticario, J. G., User Data Management and Usage Model Acquisition in an Adaptive Educational Collaborative Environment. En: Adaptive Hypermedia and Adaptive Web-Based Systems, LNCS, vol. 2347/2006, pp. 143-152, Springer Berlin / Heidelberg (2006).
- [6] Halpin, H., Robu, V., and Shepherd, H. The complex dynamics of collaborative tagging. En: Proceedings of the 16th international Conference on World Wide Web (Banff, Alberta, Canada, May 08 - 12, 2007). WWW '07, pp. 211-220. ACM, New York, NY (2007).
- [7] Moura, J., Chainho, P., and Damásio, C. V. Semantically Enabled Framework for User Centric Profile Description, Search and Match. En: Service-Oriented Computing – ICSOC 2007 Workshops: ICSOC 2007, international Workshops, Vienna, Austria, September 17, 2007, Revised Selected Papers, E. Nitto and M. Ripeanu, Eds. Lecture Notes In Computer Science, vol. 4907, pp. 326-335, Springer-Verlag, Berlin, Heidelberg (2007).
- [8] Participative Web and User-Created Content: Web 2.0, Wikis and Social Networking. SourceOCDE Science et technologies de l'information, 2007, 15. OECD Organisation for Economic Co-operation and Development (2007).
- [9] Rich, E. 1999. Users are individuals: individualizing user models. Int. J. Hum.-Comput. Stud. vol.51, no. 2, pp. 323-338, Aug. 1999.
- [10] Sato, D., Kobayashi, M., Takagi, H., and Asakawa, C. 2010. Social accessibility: the challenge of improving web accessibility through collaboration. En: Proceedings of the 2010 international Cross Disciplinary Conference on Web Accessibility (W4a) (Raleigh, North Carolina, April 26 - 27, 2010). W4A '10. ACM, New York, NY (2010).